Probability Evaluation with a Compressed Bootstrap

by

Brian K. Davis

Research Foundation of Southern California, Inc.
8861 Villa La Jolla Drive, 13595, CA 92039

(Published, 30 April, 2017)

Summary. Using a compressed bootstrap procedure substantially reduced the number of resamplings required to compute probability, in a study of cluster formation by photon trajectories during self-interference. From an initial logarithmic rate of approach to an intercept, corresponding to a hypothesized null-point, by the line fitting ($r^2 = 0.95$) ordinate cluster values at comparatively small resampling numbers, a large resampling number, determining the highly significant probability for intra-fringe clustering by reconstructed photon trajectories, was obtained. A reduction exceeding 94 per cent resulted in the resampling number.

Key words: *computed probability; resampling number; ordinate fall-off; null-point intercept*

Clustering by photon trajectories identified by Kocsis et al. (2011), within self-interference fringes, was recently evaluated from the magnitude of a bootstrap-determined probability (Davis, 2017). As a direct bootstrap computation from the observed occupancy distribution would have required a resampling number well beyond the capacity of a standard worksheet, a compressed form of the bootstrap was devised. It will be seen to place a resampling number, just large enough to contain a peak/non-peak state, with zero trajectory excess (hypothesized null-point), at the intercept of a negatively sloping line, which fit a set of ordinate mean minimum-excess values and corresponding set of comparatively small samples of increasing size. The approach adopted broadly utilizes a Kriging analysis (Krige, 1951, Lahiri, 2003; Li and Heap, 2014) to compress the standard bootstrap procedure (Efron, 1979) and compute an outcome probability.

Depicted below is the decrease in mean minimum peak-to-trough excess among photon trajectories, within a split beam interferometer (Kocsis et al., 2011), that accompanied

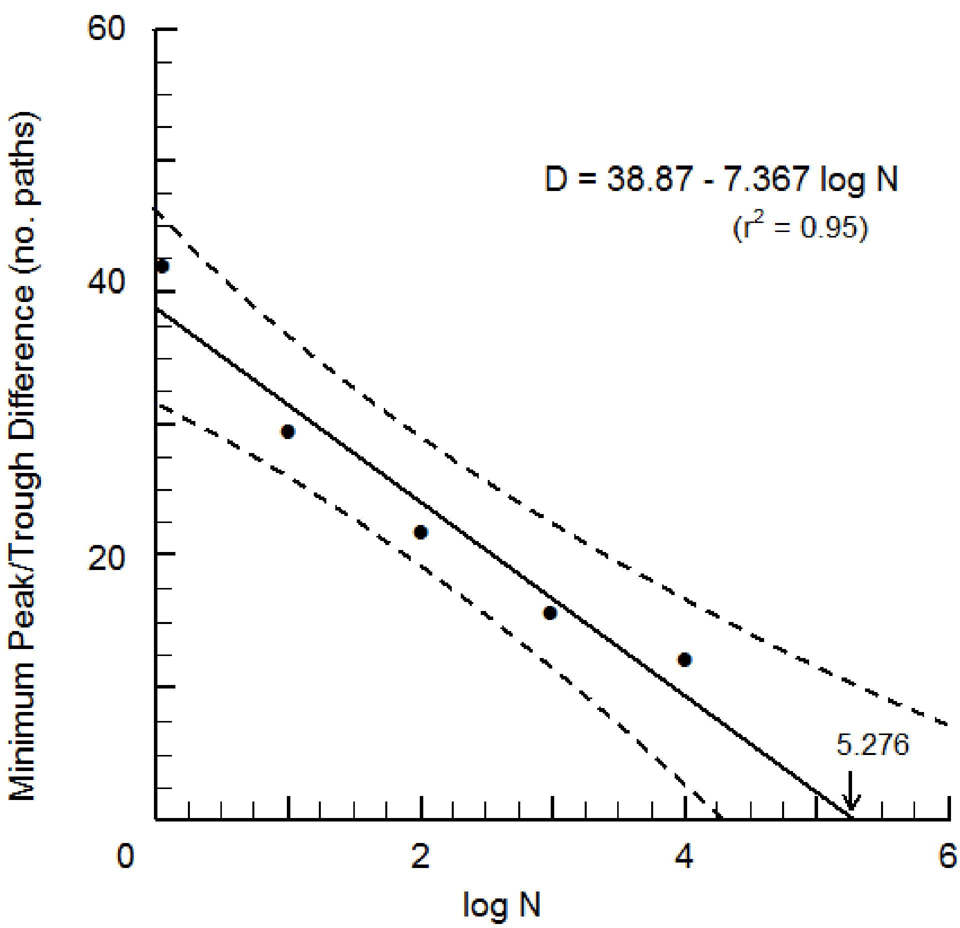

Figure. Dependence of the minimum-excess in number of peak-to-trough photon trajectories on sample size. The linear fall-off accompanying increases in log N yields an intercept at log 5.276 with 95 per cent confidence intervals of log 4.165 and log 7.649. N, sample size; r^2, coefficient of determination; and, D, mean minimum peak-to-trough difference in number of trajectories.

increases in sample size. As indicated, there were five sets of samples, each with a given number (N) of peak trajectory-excess determinations: specifically, 1, 10, 100, 1,000, and 10,000 determinations, denoted as N_i per sample, where, i = 1, 2, …, 5. Partitioning excess values in a population of 10^4 distributions, at each scale, produced the five sample sets. Formation of the population of peak trajectory-excess values resulted on repeated resampling (with replacement) of the normalized, cumulative occupancy distribution for 80 photon trajectories spread among 7 peaks and 8 troughs in a 7.7 m interferogram (see Appendix). The mean minimum-excess, for samples of a given size, was subsequently determined; with N_i determinations per sample, each sample-set contained 10,000/N_i samples.

Mean minimum peak to non-peak trajectory excess values clearly exhibit a strong dependence on sample size in a semi-log plot (see Figure). A least squares regression analysis established the coefficient of determination to be 0.95, with a statistically significant two-tail error probability of 4.35×10^{-3}, based on a Student's t-value of 7.82 at 3 degrees of freedom. Extrapolating from the line fitted to samples with 1 to 10,000 determinations, representing 10,000 to 1 resamplings of the trajectory occupancy distribution in a 7.7 m interferogram, obtained by Kocsis et al. (2011), the intercept on the abscissa is predicted to occur at log 5.276, with lower and upper 95 per cent confidence intervals of log 4.165 and log 7.649, respectively.

This places the probability of a zero peak-to-trough trajectory excess, p(δ = 0), as a 1-in-1.88 x 10^5 ($p = 5.29 \times 10^{-6}$) event – with 95 per cent numerical confidence limits of 1.46×10^4 to 4.46×10^7. A comparable estimate of $p = 4.56 \times 10^{-6}$ is obtained with a Chi-square test corrected for continuity, with a peak-to-trough distribution of 61-to-19. Since the present

estimate of the probability of the zero-excess state was based on only 10^4 resamplings of the observed photon trajectory occupancy distribution (Appendix), versus a predicted occurrence of 1 in 1.88 x 10^5, a 94.7 per cent compression of the resampling number has been achieved.

References

Davis BK 2017 Colinearity of photon path-density maxima with interference intensity peaks. http://archive.org/details/ColinearPathPeaks

Efron B 1979. Bootstrap methods: Another look at the jackknife. *Ann. Statistics* **7**, 1-26. https://dx.doi.org/10.1214/aos/1176344552

Krige DG 1951 A statistical approach to some mine valuations problems at the Witwatersrand *J. Chem. Metall. Min. Soc. S. Afr.* **52** 119-139

Lahiri SN 2003. *Resampling Methods for Dependent Data* New York: Springer-Verlag. https://dx.doi.org/10.1007/978-1-4-757-3808-2

Li J, Heap AD 2014. Spatial interpolation methods applied to the environmental sciences: A review *Environmental Modelling & Software* **53**, 173-189. https://doi.org/10.1016/j.envsoft.2013.12.008

Appendix

Table Distribution of reconstructed photon paths among peak and trough regions of a 7.7 m interferogram[a]

Regions		Photon Paths		Cum. Frequency	
no.	type	Trough	Peak	Trough	Peak
1	T-4	2		0.025	
2	**P-3**		**8**		**0.125**
3	T-3	2		0.15	
4	**P-2**		**9**		**0.2625**
5	T-2	2		0.2875	
6	**P-1**		**9**		**0.4**
7	T-1	2		0.425	
8	**P 0**		**7**		**0.5125**
9	T+1	0		0.5125	
10	**P+1**		**15**		**0.7**
11	T+2	6		0.775	
12	**P+2**		**12**		**0.925**
13	T+3	3		0.9625	
14	**P+3**		**1**		**0.975**
15	T+4	2		1	
	total:	19	**61**		

[a] **From experimental observations of Kocsis et al. (2011), and sorted by Davis (2017). Bold numbers and letters refer to photon intensity peaks (upper-half), with other numbers and letters referring to troughs.**